

Sınıflama Modelinin Performansını Değerlendirmede Yeniden Örneklem Yöntemlerinin Kullanımı

Gülhan OREKİCİ TEMEL¹, Semra ERDOĞAN^{1*}, Handan ANKARALI²

¹ Biyoistatistik ve Tıbbi Bilişim AD, Mersin Üniversitesi, Mersin, Türkiye

² Biyoistatistik ve Tıbbi Bilişim AD, Düzce Üniversitesi, Düzce, Türkiye

gulhan_orekici@hotmail.com, semraerdogann@gmail.com, hankarali@yahoo.com

(Geliş/Received: 04.05.2012; Kabul/Accepted: 25.09.2012)

Özet- Yeniden örneklem yöntemleri, her geçen gün istatistiksel araç yöntemleri olarak daha popüler hale gelmişlerdir. Literatürde, Jackknife, Bootstrap ve çapraz geçerlilik olmak üzere üç farklı yeniden örneklem yöntemi kullanılmaktadır. Yöntemler korelasyonlu yapıyı bozup, yanlılıkları ortadan kaldırmak için matematiksel model kurulumu ve kurulan modelin geçerliliğinin test edilmesinde sıklıkla kullanılmaktadır. Bu çalışmada iki farklı deneme planı uygulanmıştır. Deneme planlarında normal dağılıma sahip veriler üretilmiştir. Üretilen ham veri ve yeniden örneklem yöntemleriyle dönüştürülmüş şekli kullanılarak elde edilen veri setleri üzerinde sınıflama analizi yapılmıştır. Yeniden örneklem yöntemlerinin performansları bir simülasyon çalışması yardımıyla karşılaştırılmıştır. Verilerin oluşumu ve yeniden örneklenmesi Matlab 7.0 paket programında, istatistik analizler ise STATISTICA 7.0 paket programında yapılmıştır. Yöntemlerinin hepsinin küçük örnek genişliklerinde daha iyi sonuçlar verdiği ortaya çıkmıştır. Örneklem büyüklüğü ve geçerlilik yöntemi ne olursa olsun, Jackknife yeniden örneklem yönteminin diğer yöntemlere nazaran daha iyi bir yöntem olduğuna karar verilmiştir.

Anahtar Kelimeler- Yeniden örneklem teknikleri, simülasyon, doğruluk

Usage of Resampling Methods for Evaluating the Performance of Classification Model

Abstract- Resampling methods are becoming popular increasingly as statistical tool methods. Three different resampling methods that are Jackknife, bootstrap and cross-validation are used in the literature. Methods are often used to install of a mathematical model for eliminating bias and breaking the correlation structure and to test the validity of the established model. Two different experimental plans are applied in this study. Data with normal distribution are generated. Raw data and reproduced data which obtained by re-sampling methods analysis with classification methods. Performance of resampling methods are compared with a simulation study. Data generation and resampling are made in Matlab 7.0 program and statistical analysis is made in STATISTICA 7.0 program. All resampling methods have better results in small samples. It has been decided that Jackknife method is better than other methods regardless of sample size and validity methods.

Keywords- Resampling methods, simulation, accuracy

1. GİRİŞ

Sağlık bilimlerinde sınıflama amaçlı yapılan çalışmalar, daha çok belirli bir hastalığın tanısının konulmasında etkili olan risk faktörlerinin, istatistiksel olarak modellenerek belirlenmesi şeklinde yapılmaktadır. Bu tür sınıflama metotları ileri istatistik teknikler kullanmayı gerektirir. Bilgisayar teknolojisinin ilerlemesi, istatistik paket programlarının sınıflama ve modelleme amaçlı kullandıkları algoritmaların da yeniden örneklem yöntemlerinden herhangi birini kullanıyor olması, bu

yöntemlerin giderek daha popüler hale gelmesine neden olmuştur.

Literatürde, Jackknife, Bootstrap ve çapraz geçerlilik olmak üzere üç farklı yeniden örneklem yöntemi kullanılmaktadır. Yöntemlerin temeli permütasyon test yöntemine dayanmaktadır. Aslında permütasyon test istatistiksel anlamlılık testlerinin bir tipidir. Bu test klasik istatistikte öncü olan R. A. Fisher tarafından 1930'larda ortaya atılmıştır. Fisher, permütasyon testini, örnek ortalamaları arasındaki gözlenen farklılığın olup olmadığının belirlenmesi şeklindeki deneme planında

kullanmıştır. Bu yöntemle göre iki grubun ortalaması karşılaştırıldığında, verinin sırası ve grubu, iadesiz seçim yaklaşımı kullanılarak değiştirilip birinci, ikinci ve daha fazla permütasyonlar oluşturularak bu şekilde veriyi yeniden örnekleme [1].

Popüler sınıflama yöntemleri yeniden örnekleme yöntemlerini algoritmalarında iki şekilde kullanılmaktadır:

- (1) Model kurma
- (2) Modelin geçerliliğini test etme

Sınıflama yöntemleri model kurma algoritmalarında, model oluşumunu iki aşama da gerçekleştirmektedir. İlk aşamada yeniden örnekleme tekniklerinden herhangi birini kullanarak veri setini yeniden inşa etmektedir. İkinci aşamada ise oluşturulan yeni veri seti ile model kurulmaktadır. Birden çok değerlendiricinin ve birden çok tanı testinin diagnostik doğruluk hesaplamaları çalışmalarında (Multi Reader Multi Case Receiver Operating Curve, MRMC ROC), verilerin arasındaki korelasyonlu yapıyı bozup, yanlışlıkları ortadan kaldırmak için matematiksel model kurulumu aşamasında yeniden örnekleme yöntemlerinden Jackknife yeniden örnekleme yöntemi kullanılmaktadır. Bagging, Boosting Tree, Random Forest gibi topluluk algoritmalarına dayalı yöntemlerde ise modelin kurulumu aşamasında veri setinin yapısını bozmadan daha büyük veri seti haline getirebilmek için Bootstrap örnekleme yöntemi kullanılmaktadır [2-4].

Yeniden örnekleme yöntemlerinden çapraz geçerlilik yöntemi, hem model kurma hem de kurulan modelin geçerliliğinin test edilmesinde kullanılmaktadır. Burada en temel çapraz geçerlilik olan basit çapraz geçerlilikte, veri seti, test seti ve örneklem seti olarak ikiye ayrılmaktadır. Örneklem setinden matematiksel model kurulmakta, test setinde de modelin geçerliliği sınanmaktadır. Ayırma analizi, Lojistik regresyon, Sınıflama ve Regresyon Ağaçları (Classification and Regression Tree, CART), Çok Değişkenli Uyumlu Regresyon uzanımları (Multivariate Adaptive Regression Spline, MART) gibi sınıflama yöntemlerinin algoritmalarında modelin geçerliliğinin test edilmesi sırasında çapraz geçerlilik yöntemi kullanılmaktadır [5,6]. Bu çalışmanın amacı, sınıflama modellerinin kurulmasında ve modelin geçerliliğinin test edilmesinde sık kullanılan yeniden örnekleme yöntemlerini tanıtmak ve bir simülasyon çalışması yardımıyla sınıflama modeli oluşturmadaki performanslarını karşılaştırmaktır. Bu çalışmada sınıflama yöntemi olarak CART analizi kullanılmıştır.

2. GEREÇ VE YÖNTEMLER

2.1. Model İnşasında Kullanılan Yöntemler

2.1.1. Jackknife Yöntemi

Jackknife yeniden örnekleme yöntemi istatistiksel yanlışlıkları ortadan kaldırmak amacıyla ilk olarak 1956'da Quenouille tarafından kullanılmıştır. Daha sonra hipotez testleri ve güven aralığı oluşturmak için 1958'de Tukey tarafından da geliştirilmiştir [7,8].

Jackknife yöntemi parametrelerin güven aralıklarının tahmininde ve istatistiksel testlerin performanslarının değerlendirilmesinde büyük ölçüde kullanılan bir araçtır. Bu prosedür genellikle veri setinde aşırı uç değerlerin varlığında kullanılmaktadır. Jackknife yöntemi her aşamada bir obje ya da bireye ait veriyi dışarıda bırakarak aynı testin tekrar edilmesidir. Bu nedenle bu yöntem birini dışarıda bırak (**leave one out**) yöntemi olarak da isimlendirilmektedir. Bu şekilde bu yöntem yanlışlıkları azaltmada etkilidir [9].

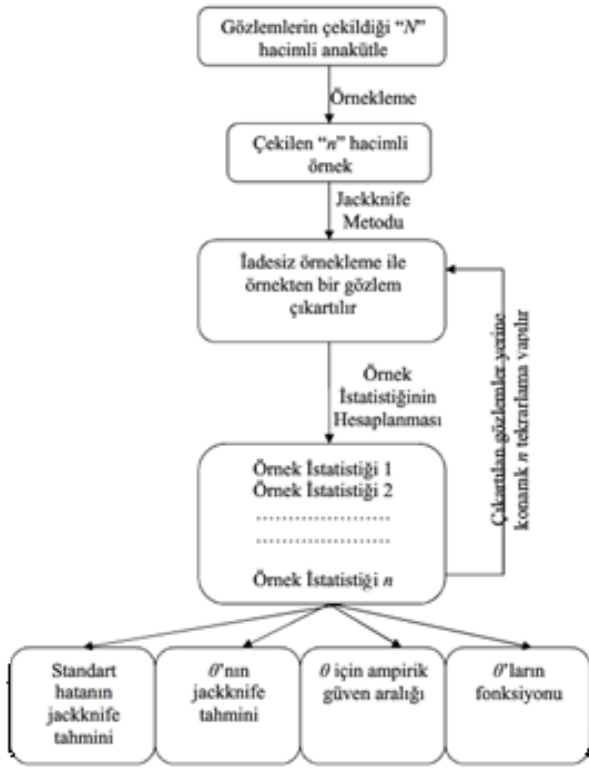
Yöntemin temel mantığı, veri setinde her bir gözlem değeri bir kez dışarıda bırakılarak geride kalan gözlemlerden örneklem istatistikleri hesaplamaya dayanmaktadır. Veri setindeki her bir gözlem dışarıda bırakılarak sahte gözlem değerleri (psueduevalues, sözde değerler) elde edilir. Sahte gözlenen değerler bir çeşit veri transformasyonu olarak bağımsız ve aynı dağılıma sahiplermiş gibi düşünülebilir. Yönteme ait şematik gösterim Şekil 1'de yer almaktadır [7-10].

Kabul edelim ki $\hat{\theta}$ parametresini tahmin etmek istiyoruz. i birey sayısını göstermek üzere $i=1,2,\dots,n$ olan bir veri setinin de genel ortalaması $\bar{\theta}$ olsun. Jackknife yönteminin her aşamasında bir obje ya da bireyin gözlem değeri silindiğinden j . veri silindiğinde örneklem ortalaması Eşitlik 1'deki gibi hesaplanmaktadır [7-9].

$$\bar{\theta}_{-j} = \frac{1}{n-1} \sum_{i \neq j}^n x_i \quad (1)$$

j . verinin gerçek gözlem değerine karşılık bu yöntemde sahte gözlem değeri hesaplanacağından bu sahte gözlem değeri de Eşitlik 2'deki gibi hesaplanmaktadır [7-9].

$$x_j = n\bar{\theta} - (n-1)\bar{\theta}_{-j} \quad (2)$$

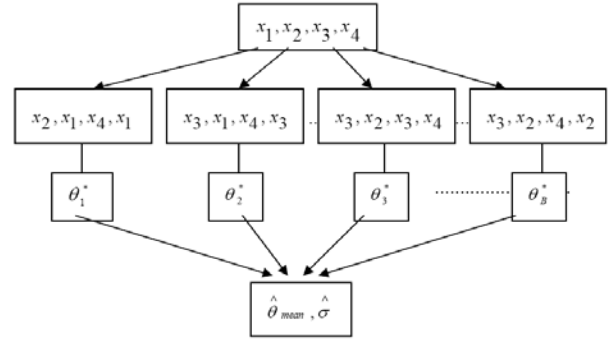


Şekil 1: Jackknife yöntemine ait şematik gösterim

2.1.2. Bootstrap Yöntemi

Yöntem ilk olarak 1979 yılında Bradley Efron tarafından Jackknife yöntemine alternatif olarak ortaya atılmıştır. 1993 yılında da Efron ve Tibshirani tarafından geliştirilmiştir [11-12]. Bootstrap metodu standart sapma, güven aralığı gibi istatistiklerde ve parametrik olmayan tahminleme problemlerinde kullanılan basit ve güvenilir bir yöntemdir. Yöntemin temel mantığı herhangi büyüklükteki mevcut veri setindeki gözlemlerin şansa bağlı olarak yer değiştirerek yeniden örneklenmesi ile yeni veri setleri oluşturulmasına dayanmaktadır. Böylece mevcut veri setinden mümkün olduğunca fazla miktarda bilgi alınabilmektedir. Yöntemde ilk olarak orijinal örneklemeden yerine koyma yöntemi ile örnekler seçilerek bir bootstrap örnekleme oluşturulmaktadır. Ardından bu şekilde çok sayıda bootstrap örneklemler oluşturularak, bu örneklemler ile tahmin edicinin güven aralığı ve testin istatistik anlamlılığı hesaplanmaktadır [13-15]. Yönteme ait şematik gösterim Şekil 2'de yer almaktadır. [16]

İstatistikte çoğu zaman incelenen parametrelerin normal dağılım varsayımına dayanılarak tahmin edilebilmesi söz konusu olurken bu yöntemde veri setinden iadeli örneklem alınarak parametreler dağılım varsayımından bağımsız olarak da tahmin edilebilmektedir. Eksik ve/veya küçük veri setleri dahi istatistiksel çıkarıma yapmak için kullanılabilir [17].



Şekil 2: Bootstrap yöntemine ait şematik gösterim

2.2. Modelin Geçerliliğini Sınama Yöntemleri

2.2.1. Yeniden Yerine Koyma Yaklaşımı

Smith, ilk olarak 1947 yılında yapmış olduğu bir çalışmada, kullanılan veri setini kurulan model sonunda, yeniden model üzerinde kullanarak hatalı sınıflama oranını tahmin etmiştir. Bu yaklaşıma yeniden yerine koyma yaklaşımı (Resubstitution estimate) denilmektedir. Kullanılan herhangi bir sınıflama modelinin geçerliliği, model kurma aşamasında kullanılan veriler ile sınıdığında, elde edilen sonuçlar yanıltıcı olmaktadır. Bu yanıltıcı durumu ortadan kaldırmak için, basit ya da çapraz geçerlilik yöntemlerinin denenerek modelin doğruluk tahminin yapılması ve geçerliliğinin sınanması gerekmektedir [18-20].

2.2.2. Çapraz Geçerlilik Yöntemleri

Basit çapraz geçerlilik (cross validation) yöntemi ilk olarak 1948 yılında Kurtz tarafından önerilmiştir. 1951 yılında Moiser bu yöntemi geliştirerek çift çapraz geçerlilik, 1982 yılında da Krus ve Fuller tarafından çoklu çapraz geçerlilik teknikleri geliştirilmiştir [21-23].

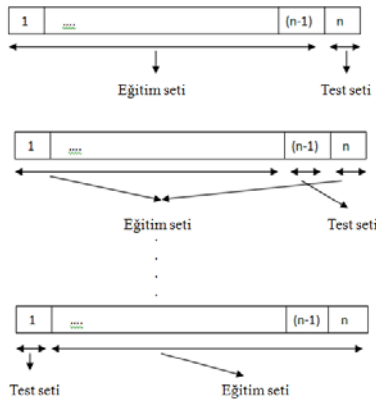
Modelin doğruluğunun test edildiği çapraz geçerlilik yönteminde veri başlangıçta test ve eğitim verisi olarak tesadüfi olarak ayrılır. Eğitim verisi modelin kurulumu (eğitimi) aşamasında kullanılır iken test verisi model kurulumunda kullanılmaz, modelin doğruluğu bu yeni veri seti üzerinde test edilir [24].

Basit çapraz geçerlilikte; veri setinin %5-%33'lük bir kısmı test verisi olarak ayrılmakta ve modelin öğrenme aşamasında bu kısım kullanılmamaktadır. Geriye kalan kısım üzerinde ise model kurulmakta ve gerçek değerler ile tahmin değerleri karşılaştırılarak modelin doğruluğu hesaplanmaktadır.

Çift çapraz geçerlilik ve çoklu çapraz geçerlilik, veri setinin sınırlı sayıda olması durumunda kullanılmaktadır. Bu yöntemler, modelleme aşamasında bütün veri setinin kullanımına imkân sağlamaktadır. Veri seti, bir bölümü eğitim verisi diğer bölümü test verisi olmak üzere tesadüfi olarak iki eşit parçaya ayrılır ve bu şekilde bir doğruluk

hesaplaması yapılır. Ardından test verisi ile eğitim verisi yer değiştirilerek yeniden bir doğruluk hesaplaması yapılır ve doğruluk değerlerinin ortalaması alınarak modelin doğruluk oranı hesaplanır. Bu algoritma çift çapraz geçerlilik olarak tanımlanmaktadır.

Çoklu çapraz geçerlilik yöntemi ise, her verinin bir kez eğitim bir kez de test verisi olarak kullanabileceği geçerlilik yöntemlerinin genellenmiş halidir. Bu yöntemde, veri n adet eşit gruba ayrılır. Bir grup test olarak, geriye kalan $(n-1)$ grup ise eğitim verisi olarak kullanılmaktadır ve bu durum n kez tekrar edilir. Yine doğruluk değerlerinin ortalaması alınarak modelin doğruluk oranı belirlenir [25]. Yönteme ait şematik gösterim Şekil 3'de yer almaktadır.



Şekil 3. Çoklu Çapraz geçerlilik yöntemine ait şematik gösterim

2. 3. Sınıflama Ağaçları

Sınıflama Ağaçları (CART) kategorik yapıdaki bağımlı değişkenin alacağı değerleri tahmin etmeye yönelik geliştirilen parametrik olmayan bir veri madenciliği yöntemidir. Modelde değişkenlerin türü (sürekli, kategorik, sıralı ya da bunların karışımı) konusunda herhangi bir varsayım gerektirmez ve sınırlama getirmez. Hem bağımlı hem de bağımsız değişkenler için kayıp veya eksik değerler ile aşırı uç değerlerden etkilenmeyen bir metottur.

Buna karşın CART tekniğinin en önemli sınırlılığı sonuçların bir olasılık modeline dayanmıyor olmasıdır. Eğitim seti adı verilen verinin bir bölümü ile CART ağacı inşa edilir. Test verisi ile de ağacın geçerliliği test edilir. Ağaç inşası kök düğümden başlayarak devam eden ve her düğümden o düğüme ait deney ünitelerine uygulanan basit sorulardan alınan evet/hayır cevaplarına göre oluşan yollar içerir. Her düğümden uygulanan bu sorulara ayırac denir. Oluşan ağaçlarda homojen olmayan düğümlere çocuk düğümleri, homojen düğümlere ise terminal düğümler adı verilir ve hesaplamalar sonucunda terminal düğümler içindeki gözlem sayısının yoğunluğuna göre 1.grup veya 2.grup olarak adlandırılır ve bu düğümler yorumlanır. Bu yöntemde, doğru sınıflamanın başarısını ölçmede yeniden yerine koyma yaklaşımı, test örneği tahmini ve çapraz

geçerlilik yöntemleri birbirine alternatif olarak kullanılabilir [5].

2.4. Simülasyon Çalışması

Bu çalışmada iki farklı deneme planı tasarlanmıştır. Bu deneme planları için simülasyonla üretilen verilerin özellikleri Tablo 1'de verilmiştir. Her iki deneme planında da amaç, CART yöntemi yardımıyla, iki farklı popülasyondan alınan iki grubu birbirinden ayırmaya çalışmaktır. Birçok sınıflama yöntemi, örneklem genişliği, değişken sayısı ve değişkenler arası ilişkilerden etkilenmektedir. Bu nedenle 1.deneme planında bir bağımsız değişken ve örnek genişliklerinin de sırasıyla 30 ve 100 olduğu durum dikkate alınmış, 2. deneme planında ise aralarında korelasyon bulunmayan 3 bağımsız değişken olduğu varsayılmıştır. Tablo 1' de görüldüğü üzere 1. grup ortalaması 0, ikinci grup ise ortalaması 1 standart sapma olan ve standart normal dağılım gösteren iki farklı popülasyondan alınmıştır. Ortalamaları farklı olan popülasyonlardan örnek alınarak gruplar oluşturulduğu için bu grupların sınıflama yöntemleri ile başarılı bir şekilde ayrıştırılması beklenir. Grupların sınıflandırılmasında CART yöntemi kullanılmıştır. CART yöntemi ile sınıflama, orijinal veri, Jackknife ve Bootstrap dönüşmüş hali için 10-katlı çapraz geçerlilik algoritması kullanılarak ve bu algoritma kullanılmadan (yeniden yerine koyma yaklaşımı) yapılmıştır. Sınıflamada birinci deneme planında bir özellik, ikinci deneme planında ise üç özellik kullanılmıştır. Her bir deneme kombinasyonu 1000 kez tekrarlanmış olup performans ölçüsü olarak doğru sınıflama başarılarının ortalaması alınmıştır.

Verilerin üretilmesinde Matlab 7.0 paket programı, veri analizinde ise STATISTICA 7.0 paket programı kullanılmıştır.

3. BULGULAR

Örneklem büyüklükleri 30 ve 100 için, oluşturulan orijinal veri, Jackknife ve Bootstrap dönüşmüş hali için 10-katlı çapraz geçerlilik algoritması kullanılarak ve kullanılmadan, tek bir sürekli özelliğin var olduğu (Deneme Planı 1) durumda elde edilen sonuçlar Tablo 2'de, üç tane sürekli özelliğin var olduğu (Deneme Planı 2) durumdaki sonuçlar ise Tablo 3'de verilmektedir.

Tablo 2 incelendiğinde, yeniden yerine koyma yaklaşımına göre, popülasyonlardan çekilen örneklem büyüklüklerinin 30 olduğu durumda oluşturulan CART analizi sonucunda doğru sınıflama başarısı % 91.7 iken örneklem büyüklüğü 100 olduğu durumda % 81.4 olarak elde edilmiştir. Orijinal veri seti jackknife yeniden örnekleme yöntemine göre yeniden üretildiğinde ve bu yeni veri setine CART analizi uygulandığında örneklem büyüklüğü 30 ve 100 için % 100 doğruluk başarısına ulaşılmıştır. Bootstrap yeniden örnekleme yöntemine göre yeniden üretildiğinde örneklem büyüklüğü 30 için sınıflama başarısı % 94.0 iken örneklem büyüklüğü 100 için % 64.5 olarak elde edilmiştir.

10-katlı çapraz geçerlilik sonuçları incelendiğinde, örneklem büyüklüklerinin 30 olduğu durumda oluşturulan CART analizi sonucunda doğru sınıflama başarısı % 79.8 iken örneklem büyüklüğü 100 olduğu durumda % 72.0 olarak elde edilmiştir. Orijinal veri seti jackknife yeniden örnekleme yöntemine göre yeniden üretildiğinde ve bu yeni veri setine ait CART analizi uygulandığında örneklem büyüklüğü 30 ve 100 için % 100 doğruluk başarısına ulaşılmıştır. Bootstrap yeniden örnekleme yöntemine göre yeniden üretildiğinde örneklem büyüklüğü 30 için sınıflama başarısı % 86.0 iken örneklem büyüklüğü 100 için % 80.1 olarak elde edilmiştir.

Tablo 3 incelendiğinde, yeniden yerine koyma yaklaşımına göre, örneklem büyüklüklerinin 30 olduğu durumda oluşturulan CART analizi sonucunda doğru sınıflama başarısı % 94.5 iken örneklem büyüklüğü 100 olduğu durumda % 87.0 olarak hesaplanmıştır.

Orijinal veri seti Jackknife yeniden örnekleme yöntemine göre yeniden üretildiğinde ve bu yeni veri setine CART analizi uygulandığında örneklem büyüklüğü 30 ve 100 için % 100 doğruluk başarısına ulaşılmıştır. Bootstrap yeniden örnekleme yöntemine göre yeniden üretildiğinde örneklem büyüklüğü 30 için sınıflama başarısı % 96.2 iken örneklem büyüklüğü 100 için % 88.5 olarak elde edilmiştir.

10-katlı çapraz geçerlilik sonuçları incelendiğinde, örneklem büyüklüklerinin 30 olduğu durumda oluşturulan CART analizi sonucunda doğru sınıflama başarısı % 84.0 iken örneklem büyüklüğü 100 olduğu durumda % 81.0 olarak hesaplanmıştır. Orijinal veri seti Jackknife yeniden örnekleme yöntemine göre yeniden üretildiğinde ve bu yeni veri setine CART analizi uygulandığında örneklem büyüklüğü 30 ve 100 için % 100 doğruluk başarısına ulaşılmıştır. Bootstrap yeniden örnekleme yöntemine göre yeniden üretildiğinde örneklem büyüklüğü 30 için sınıflama başarısı % 88.2 iken örneklem büyüklüğü 100 için % 84.3 olarak elde edilmiştir.

Tablo 1. Deneme planlarına ait oluşturulan verilerin özellikleri

Populasyonlardan çekilen örnek genişlikleri (n_i)	Deneme Planı 1		Deneme Planı 2	
	Bir özellik		Üç özellik	
	Populasyon 1	Populasyon 2	Populasyon 1	Populasyon 2
30	$N \sim (0,1)$	$N \sim (1,1)$	$N \sim (0,1)$ Her bir değişkenin arasındaki korelasyon sıfır ($r_{xyz}=0$)	$N \sim (1,1)$ Her bir değişkenin arasındaki korelasyon sıfır ($r_{xyz}=0$)
100	$N \sim (0,1)$	$N \sim (1,1)$	$N \sim (0,1)$ Her bir değişkenin arasındaki korelasyon sıfır ($r_{xyz}=0$)	$N \sim (1,1)$ Her bir değişkenin arasındaki korelasyon sıfır ($r_{xyz}=0$)

Tablo 2. Deneme planı 1 için sonuçlar

		n=30		n=100	
		Min-max	Ort \pm SD	Min-max	Ort \pm SD
Orijinal Veri	Yeniden Yerine Koyma Yaklaşımı	0.850-0.980	0.917 \pm 0.030	0.740-0.890	0.814 \pm 0.026
	10-katlı Çapraz Geçerlilik	0.50-0.950	0.798 \pm 0.090	0.600-0.820	0.720 \pm 0.032
Jackknife	Yeniden Yerine Koyma Yaklaşımı	1.000-1.000	1.000 \pm 0.000	1.000-1.000	1.000 \pm 0.000
	10-katlı Çapraz Geçerlilik	1.000-1.000	1.000 \pm 0.000	1.000-1.000	1.000 \pm 0.000
Bootstrap	Yeniden Yerine Koyma Yaklaşımı	0.867-1.000	0.940 \pm 0.030	0.790-0.935	0.645 \pm 0.91
	10-katlı Çapraz Geçerlilik	0.667-0.983	0.860 \pm 0.080	0.856 - 0.027	0.801 \pm 0.052

Tablo 3. Deneme planı 2 için sonuçlar

		n=30		n=100	
		Min-max	Ort \pm SD	Min-max	Ort \pm SD
Orijinal Veri	Yeniden Yerine Koyma Yaklaşımı	0.880-1.000	0.945 \pm 0.023	0.81-0.92	0.87 \pm 0.02
	10-katlı Çapraz Geçerlilik	0.500-0.970	0.840 \pm 0.068	0.71-0.90	0.81 \pm 0.04
Jackknife	Yeniden Yerine Koyma Yaklaşımı	1.000-1.000	1.000 \pm 0.000	1.000-1.000	1.000 \pm 0.000
	10-katlı Çapraz Geçerlilik	1.000-1.000	1.000 \pm 0.000	1.000-1.000	1.000 \pm 0.000
Bootstrap	Yeniden Yerine Koyma Yaklaşımı	0.880-1.000	0.962 \pm 0.019	0.84-0.94	0.885 \pm 0.020
	10-katlı Çapraz Geçerlilik	0.730-1.000	0.882 \pm 0.063	0.72-0.93	0.843 \pm 0.042

4. TARTIŞMA VE SONUÇ

Kullanılan sınıflama modelinin geçerliliğini, model sonunda yanlış sınıflanmış bireylerin oranını elde ederek ölçmek çok yaygın kullanılan bir yöntemdir. Literatürde sınıflama yapılırken yeniden yerine koyma yaklaşımı ve çapraz geçerlilik yaklaşımları kullanılmaktadır. Kullanılan herhangi bir sınıflama modelinin geçerliliği, model kurma aşamasında kullanılan aynı veriler (yeniden yerine koyma tekniği) ile sınıflandığında fazla iyimser olmakta ve yanıltıcı sonuçlar vermektedir. Schiavo ve ark. çalışmalarında, sınıflama çalışmalarında hatalı sınıflama oranlarının hesaplanırken yeniden yerine koyma tekniğinin gerçek hatayı doğru olarak ortaya koymadığına değinmişlerdir. Ayrıca, bu yanlılığı ortadan kaldırmak için çapraz geçerlilik, jackknife ve bootstrap yöntemlerinin daha iyi performans gösterdiklerini ortaya koymuşlardır [19].

Braga-Neto ve ark. hatalı sınıflama oranı hesaplamalarında iki genel yöntem olan çapraz geçerlilik ve yeniden yerine koyma yaklaşımları üzerine çalışmalar yapmışlardır. Yanlılığı ortadan kaldırdığı için çapraz geçerlilik yöntemini önermelerine karşın, küçük örnek genişliklerinde yeniden yerine koyma yaklaşımının daha iyi sonuç verdiğini belirtmişlerdir [20].

Chernick ve ark. sınıflama başarısını artırmak için yeniden örnekleme tekniklerinin performansları üzerinde bir çalışma yapmışlar ve bu çalışma sonucunda çapraz geçerlilik yöntemini önermişlerdir [26]. Verbyla ve ark. ise üzerinde durulan özelliğin varlık ya da yokluğunun tamamen tesadüfi olarak oluşturulduğu (ayırma gücü %50) örneklem büyüklüğünün 30 olarak alındığı ve 10 bağımsız değişkenin var olduğu iki grup üzerinde bir simülasyon çalışması yapmışlardır. Burada sınıflama yöntemi olarak ayırma analizini kullanmışlardır. Mevcut verileri yeniden örnekleme yöntemlerine göre yeniden elde etmişler ve modelin doğru sınıflama başarısına bakmışlardır. Yeniden yerine koyma yaklaşımı kullanılarak, modelin doğru sınıflama başarısına bakıldığında, modelin doğruluğunun yanlı tahmin

edildiğini ortaya koymuşlardır. Yeniden yerine koyma yaklaşımı dışındaki diğer yöntemlerin neredeyse yansız tahminlerde buldukları sonucuna varmışlardır. Ayrıca, bu çalışma ile Bootstrap örnekleme yönteminin en iyi sonucu verdiğini gözlenmiştir [27].

Bu çalışmada ise sınıflama oranı, hem yeniden yerine koyma yaklaşımı hem de çapraz geçerlilik yöntemini kullanarak hesaplanmıştır. Sınıflama oranının yeniden yerine koyma yaklaşımı kullanılarak hesaplandığında bütün yöntemlerde doğru sınıflama oranı yüksek çıkmıştır. Bu sonucun yüksek çıkmasının da yeniden yerine koyma yaklaşımının neden olduğu yanlılıktan kaynaklandığı düşünülmektedir.

Simülasyon sonuçları incelendiğinde, küçük örnek genişliklerinde yeniden örnekleme yöntemlerinin büyük örnek genişliklerine nazaran daha iyi sonuçlar verdiği görülür. Küçük örneklerde sınıflama başarısının daha iyi olması, büyük örneklerde bu yöntemlerin daha başarısız olduğu anlamına gelmez. Çünkü örnek genişliği arttıkça varyasyon artacağı için bütün sınıflama yöntemlerinin sınıflama başarısı biraz düşecektir. Özellikle orta veya küçük örneklerde sınıflama hatasının daha az olması beklenir. Bağımsız değişken sayısının artması ise sınıflama başarısını artıran bir diğer etkidir. Buna karşın CART yöntemi, örneklem büyüklüğünden, uç değerlerden ve değişkenler arasındaki çok yüksek ilişkilerden çok etkilenmeyen bir yöntemdir.

Sonuç olarak, bu çalışmada dikkate alınan örneklem genişliklerinde Jackknife yeniden örnekleme yönteminin daha başarılı sonuçlar verdiğini görülmüştür. Ayrıca Jackknife yönteminin doğası gereği veri seti üretimi aşamasında aşırı uç değer üretmemesi ya da var olan uç değerleri elimine etmesi yöntemin bir başarısı olarak düşünülebilir.

KAYNAKLAR

- [1] R.A. Fisher, **The Design of Experiment**, Hafner Press, New York, A.B.D., 1960
- [2] D.D. Dorfman, K.S. Berbaum, R.V. Length. "Multireader Multicase Receiver Operating Characteristic Methodology: A Bootstrap Analysis", *Academic Radiology*, 2(7), 626-633,1995.
- [3] M. Skurichina, R. Duin, "Bagging, Boosting and The Random Subspace Method for Linear Classifiers", *Pattern Analysis & Applications*, 5:121-135,2002.
- [4] G. Death, "Boosted Trees for Ecological Modelling and Prediction", *Ecological Society of America*, 88(1), 243-251,2007.
- [5] L. Brieman, J.H. Friedman, R.A. Olshen, C.J. Stone, **Classification and Regression Trees**, Chapman&Hall/Crc, U.S.A 1984.
- [6] J.H. Friedman, "Multivariate Adaptive Regression Splines (With Discussion)", *Annals of Statistics*, 19 (March):1-141,1991.
- [7] M.H. Quenouille, "Notes on Bias in Estimation", *Biometrika*, 43: 353-360,1956.
- [8] J.W. Tukey, "Bias and Confidence in Not-Quite Large Samples (Abstract)", *The Annals of Mathematical Statistics*, 29(2), 614,1958.
- [9] C.H. Yu, "Resampling Methods: Concepts, Applications and Justification", *Practical Assessment, Research & Evaluation*, 8(19),2003.
- [10] M.K. Beşer, "**Zaman Serilerinde Bootstrap Çözümlemeleri ve Türkiye'de Tanzi Etkisine Uygulaması**", (Doktora tezi), Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, 2006.
- [11] B. Efron, "Bootstrap Methods: Another Look at The Jackknife", *The Annals of Statistics*,7(1): 1-26, 1979.
- [12] B. Efron, R. Tibshirani, **An Introduction to The Bootstrap**, Chapman&Hall, New York, A.B.D.,1993.
- [13] B. Efron, "Censored Data and Bootstrap", *Journal of the Acoustical Society of America*, 76 (374), 312-319,1981.
- [14] I. Peterson, "Pick a Sample", *Science News*, 140: 56-57, 1991.
- [15] N. Hamajima, H. Yuasa, K Matsuo, "Methods for Statistical Inferences", *Biotherapy* 13(6), 739-44,1999.
- [16] Ç. Takma , H. Atıl, "Bootstrap Metodu ve Uygulanışı Üzerine Bir Çalışma 1. Olasılık ve Bootstrap Metodu", *Ege Üniversitesi Ziraat Fakültesi Dergisi*, 40 (3):89-96, 2003.
- [17] R. Wehrens, H. Putter, L.M.C. Buydens, "The Bootstrap: A Tutorial", *Chemometrics and Intelligent Laboratory Systems*, 54, 35-52,2000.
- [18] C.A.B. Smith, "Some Examples of Discrimination", *The Annals of Eugenics*,13,272-282, 1947.
- [19] R.A. Schiavo, D.J. Hand, "Ten More Years Rate Research", *International Statistical Review*, 68(3),295-310, 2000.
- [20] U. Braga- Neto, R. Hashimoto, E.R. Dougherty, D.V. Nguyen, R. Carrol, "Is Cross-Validation Better than Resubstitution for Ranking Genes?", *Bioinformatics*, 20(2),253-258, 2004.
- [21] A.K. Kurtz, "A Research Test of The Rorschach Test", *Personnel Psychology*, 1,41-53, 1948.
- [22] C.I. Mosier, "Problems and Designs of Cross-Validation", *Educational and Psychological Measurement*,11, 5-11, 1951.
- [23] D.J. Krus, E.A. Fuller, "Computer-Assisted Multicross-Validation in Regression Analysis", *Educational and Psychological Measurement*, 42, 187-193, 1982.
- [24] C.M. Bishop, **Neural Networks for Pattern Recognition**, Clarendon Press, Oxford, 1995.
- [25] T. Hastie, R. Tibshirani, J. Friedman, **The Elements of Statistical Learning Data Mining. Inference and Prediction**, Springer, Canada, 2003.
- [26] M.R. Chernick, VK Murthy, CD Nealy, "Application of Bootstrap and Other Resampling Techniques: Evaluation of Classifier Performance", *Pattern Recognition Letters*, 3, 167-178, 1985.
- [27] D.L. Verbyla, J.A. Litvaitis, Resampling Methods for Evaluating Classification Accuracy of Wildlife Habitat Models, *Environmental Management*, 13(6), 783-787, 1989.